Implicit Solvent Models: Combining an Analytical Formulation of Continuum Electrostatics with Simple Models of the Hydrophobic Effect

FRANÇOIS WAGNER, THOMAS SIMONSON

Laboratoire de Biologie Structurale, Institut de Génétique et Biologie Moléculaire et Cellulaire (C.N.R.S.), 1 Rue Laurent Fries, 67404 Strasbourg–Illkirch, France

Received 31 March 1998; accepted 4 September 1998

ABSTRACT: The recent development of approximate analytical formulations of continuum electrostatics opens the possibility of efficient and accurate implicit solvent models for biomolecular simulations. One such formulation (ACE, Schaefer & Karplus, *J. Phys. Chem.*, 1996, 100:1578) is used to compute the electrostatic contribution to solvation and conformational free energies of a set of small solutes and three proteins. Results are compared to finite-difference solutions of the Poisson equation (FDPB) and explicit solvent simulations and experimental data where available. Small molecule solvation free energies agree with FDPB within 1–1.5 kcal/mol, which is comparable to differences in FDPB due to different surface treatments or different force field parameterizations. Side chain conformation free energies of aspartate and asparagine are in qualitative agreement with explicit solvent simulations, while 74 conformations of a surface loop in the protein Ras are accurately ranked compared to FDPB. Preliminary results for solvation free energies of small alkane and polar solutes suggest that a recent Gaussian model could be used in combination with analytical continuum electrostatics to treat nonpolar interactions. © 1999 John Wiley & Sons, Inc. J Comput Chem 20: 322–335, 1999

Keywords: implicit solvent models; continuum electrostatics; hydrophobic effect

Correspondence to: T. Simonson; e-mail: simonson@igbmc.u-strasbg.fr

Contract/grant sponsor: Centre National de la Recherche Scientifique

CCC 0192-8651 / 99 / 030322-14

Introduction

Computer simulations are a powerful tool for the study of biological molecules in solution. Much of their cost and complexity arises from the detailed treatment of aqueous solvent surrounding the solute of interest. The development of accurate implicit models of aqueous solvation would alleviate much of this cost, increasing the solute size and the time scale amenable to simulation.[1] A promising approach has been to treat the solute–solvent electrostatic interaction by continuum electrostatics[2–5] and the remaining sum of interactions as a simple empirical function of solute surface area or geometry.[6–8] Thus, vapor to water transfer free energies of small organic molecules can be calculated with reasonable accuracy as a sum of a continuum electrostatic free energy and a nonpolar term proportional to the solute accessible surface area.[9,10] The continuum electrostatic free energy and forces are usually obtained by numerically solving the Poisson–Boltzmann equation.[11] While this is much faster than simulating the explicit solvent, it still entails a significant computational cost. A set of related, analytical approximations to this free energy have been proposed recently, which are expected to be much more efficient.[12–18] In these approaches the solvation of individual solute charges is calculated as if the charge were in a homogeneous solvent environment (i.e., neglecting other solute groups) and the Coulombic interactions between solute charges are reduced by an empirical screening function, the ''generalized Born'' function of Still and coworkers.[12] We focus here on one implementation of these ideas, termed analytical continuum electrostatics (ACE) by Schaefer and Karplus (SK).[16] The main goal of this article is to compare different parameterizations of this method and assess its accuracy for solvation and conformational free energies of small organic solutes and proteins.

The remaining, nonelectrostatic interactions between solute and solvent are more difficult to model simply. Thus, while a single microscopic surface tension of $\gamma \approx 6$ cal/mol/Å^2 is sufficient to describe the solvation free energies of linear and moderately branched saturated alkanes (i.e., $\Delta G \approx \gamma A$, where A is the solute accessible surface area), the same functional form fails for cyclic saturated alkanes.[10] A different approach was proposed recently, which expresses the solvation energy of nonpolar and polar compounds as a pairwise sum of Gaussian interatomic contributions.[19,20] This approach is used below to complement the continuum electrostatic term and describe nonpolar interactions.

The ACE approach is applied to a set of small charged and polar organic solutes ranging in size from acetate to methyl-indole. The ACE estimate of the solvation free energy is compared to the finite-difference numerical solution of the Poisson–Boltzmann equation (FDPB). Calculations are done with two force fields commonly used in molecular dynamics simulations (OPLS[21] and CHARMM22),[22] and different schemes for partitioning the molecular volume among atoms and defining the solute–solvent boundary are compared. Calculations of the electrostatic component of the solvation free energy are also reported for the proteins bovine pancreatic trypsin inhibitor (BPTI), crambin, and Ras. The CHARMM19 force field is used[23] for comparison with calculations of SK on BPTI.[16] To test the ability of ACE to discriminate between multiple conformers of a given solute, conformational energies are calculated for all side chain conformations of the two amino acids aspartate (Asp) and asparagine (Asn) and compared to existing molecular dynamics and free energy simulations in explicit solvent.[24] Finally, conformational energies are calculated for 74 conformations of a flexible surface loop of the Ras protein (residues 30–37 in the effector domain of the protein) and compared to FDPB calculations. For the latter system, ACE was 700 times faster than FDPB.

To evaluate the Gaussian solvation model for nonpolar compounds, a data set of experimental solvation free energies of 23 saturated alkanes was used. Adjustment of two model parameters led to excellent agreement between calculated and experimental values [0.19 kcal/mol rms error], including those for seven cyclic alkanes. Application to 29 polar compounds is also reported. In this case electrostatic interactions were treated with FDPB and the Gaussian contribution was used to model the remaining, primarily hydrophobic interactions. After adjustment of three model parameters, agreement for the solvation free energies was poorer than for the alkanes (1.2 kcal/mol rms error) but slightly superior to earlier calculations with an accessible surface area model.[10] This suggests that a hybrid model combining ACE or FDPB with a Gaussian treatment of nonpolar interactions may be a useful approximation. The performance of such a hybrid model for large solutes such as proteins remains to be tested.

Methods

ACE IMPLEMENTATION

In continuum electrostatics a solute is viewed as a set of (fractional) atomic charges in a cavity delimited by the solute surface, which is embedded in a high dielectric solvent medium.[2] The solvation free energy can be expressed as a sum of the interaction energies of each solute charge with the solvent (their ''self-energies'') and the interaction energies between all pairs of solute charges. In the ACE approach the interaction energy E_{ij}^{int} between charges q_i and q_j is approximated by the generalized Born expression[12]:

$$E_{ij}^{\text{int}} = \frac{q_i q_j}{r_{ij}} - \frac{(1 - 1/\varepsilon_w) q_i q_j}{\left(r_{ij}^2 + b_i b_j \exp\left[-r_{ij}^2/4 b_i b_j\right]\right)^{1/2}}, \quad (1)$$

where r_{ij} is the distance between the charges, ε_w is the solvent dielectric constant, b_i and b_j are solvation radii of charges i and j, respectively:

$$b_i = -\frac{(1 - 1/\varepsilon_w) q_i^2}{2\,\Delta E_i^{\text{self}}}, \quad (2)$$

and ΔE_i^{self} is the self-energy of charge i given below. The solute dielectric constant is taken to be one. This and the use of fixed atomic partial charges implies that charge redistribution within the solute in response to solvation is not included in the model.

By introducing Gaussian atomic volumes, SK were able to express the self-energy of a charge i as a sum over pairs of solute atoms[16]:

$$\Delta E_i^{\text{self}} = -\frac{(1 - 1/\varepsilon_w) q_i^2}{2 R_i} + \sum_{k \neq i} E_{ik}^{\text{self}}, \quad (3)$$

where R_i is an atomic radius to be determined (but close to the van der Waals radius) and E_{ik}^{self} is approximated by the Ansatz

$$E_{ik}^{\text{self}} = \frac{(1 - 1/\varepsilon_w) q_i^2}{\omega_{ik}} \exp(-r_{ik}^2/\sigma_{ik}^2) + \frac{(1 - 1/\varepsilon_w) q_i^2 V_k}{8\pi} \left(\frac{r_{ik}^3}{r_{ik}^4 + \mu_{ik}^4}\right)^4. \quad (4)$$

Here ω_{ik}, σ_{ik}, and μ_{ik} are simple functions of the atomic volume V_k, the atomic radii R_i, R_k, and a ''smoothing'' parameter α that determines the width of the atomic Gaussian distributions, respectively.

The adjustable parameters of the model are the atomic charges q_i, radii R_i, and volumes V_i, and the smoothing parameter α. Ionic strength is not included in the model. In the following, we make use of three standard sets of charges and radii (OPLS, CHARMM19, CHARMM22), and volumes are either calculated using Voronoi polyhedra (with several different algorithms[25]) or assigned values from the library of ref. 16. The ACE solvation energy was implemented in the program X-PLOR.[26]

ATOMIC VOLUME ASSIGNMENTS FOR ACE

For small solutes (following SK[16]) atomic volumes for non-hydrogen atoms were obtained from atomic Voronoi polyhedra using the program VOLUME of Richards.[27] Polyhedra are constructed from planes normal to the lines connecting each pair of atoms, either bisecting the line or slightly offset toward the smaller atom of the pair, to assign a greater proportion of the volume to the larger atom in a way that depends on the relative atomic radii.[25,27] These methods will be referred to as the bisecting plane (BP) and the offset plane (OP) methods, respectively. Hydrogen atoms were assigned either a constant volume or the volume of their van der Waals sphere.

For the proteins a library of atomic volumes for ACE[16] was used for most calculations. Voronoi volumes were used for comparison.

FDPB CALCULATIONS

Calculations were done with the program UHBD,[28,29] which maps the system onto a 3-dimensional cubic grid and performs several iterative, ''focussing'' calculations with finer and finer grids. For small solutes the total grid length was 55 Å and the final grid spacing was 0.2 Å. For proteins the total grid length was 108 Å and the final grid spacing was 0.4 or 0.5 Å. The solute–solvent boundary is defined as the molecular surface[25], calculated with a 1.4-Å solvent probe sphere and using atomic radii equal to the minimum of the van der Waals potential for the atom type in the force field used (e.g., CHARMM19 or CHARMM22). Dielectric boundary smoothing was used,[28,29] which slightly blurs the boundary. The solute dielectric constant was one in all cases. Ionic

strength was set to zero. Results were averaged over eight slightly different positions of the grid relative to the solute. Calculations on small solutes were done with the CHARMM22 force field and a modified version of this force field where the atomic radii were specifically optimized for Poisson calculations.[30] Earlier calculations on small solutes in similar conditions but with the OPLS force field are also available for comparison.[10]

GAUSSIAN NONPOLAR SOLVATION MODEL

This model was originally developed as an empirical implicit solvent model for polar and nonpolar solutes.[19,20] Here we use it to describe only nonelectrostatic solute–solvent interactions (i.e., we view it as a nonpolar energy term, which is complementary to the continuum electrostatic term above). We test it for saturated alkanes where electrostatic interactions are negligible and for a set of small polar compounds where the electrostatic interactions are treated by FDPB. The contribution G_i of an atom i to the nonpolar energy term has the form

$$G_i = G_i^{\mathrm{ref}} - \sum_j f_i(r_{ij})V_j, \qquad (5)$$

where the sum is over all other solute atoms j, G_i^{ref} is a constant reference energy, V_j is the volume of atom j, r_{ij} is the distance between atoms i and j, f_i is given by

$$f_i(r) = \frac{G_i^{\mathrm{free}}}{2\pi^{3/2}\lambda_i r^2}\exp\left[-(r - R_i^{\mathrm{min}})^2/\lambda_i^2\right], \qquad (6)$$

G_i^{free} is another constant energy, R_i^{min} is the radius of atom i, and λ_i is an adjustable constant.

Initial model parameters for alkane atoms were taken from Lazaridis and Karplus[19,20]; these include a single $\lambda_i = \lambda$ value for all hydrocarbon atom types. This λ was then adjusted along with a single overall scaling factor to reproduce experimental solvation free energies of 23 alkanes.

The model was then applied to 29 polar solutes. Atomic radii and volumes were taken from Lazaridis and Karplus.[19,20] A single adjustable value of G^{ref} and G^{free} was introduced, which was applicable to all non-alkane atoms, and an adjustable scaling factor was applied to the G^{ref} and G^{free} of all alkane atoms (in these polar solutes).

CONFORMATIONAL AND SOLVATION FREE ENERGIES OF Asp AND Asn

Adiabatic (χ_1, χ_2) energy maps for aspartate and asparagine (with their backbones in zwitterionic form) were calculated using the CHARMM22 force field for the intramolecular energies and treating the solvent implicitly, by either setting the dielectric constant to 20 for the intramolecular Coulombic interactions or using ACE. The ACE results for differences between conformations can be viewed as approximate free energy differences, which is consistent with the usual methods for calculating free energies with FDPB (see refs. 9, 10). Free energy differences are also reported using ACE and FDPB for the alchemical transformation of Asp into Asn in solution for which free energy perturbation calculations with explicit solvent are available.[24] The ACE and FDPB calculations make use of the following transformation pathway: first Asp is removed from the solvent to a vacuum, next the mutation is performed in a vacuum, and finally Asn is immersed in the solvent. Free energy perturbation results for the *in vacuo* transformation using molecular dynamics were obtained elsewhere (G. Archontis, T. Simonson, and M. Karplus, unpublished data).

Results

POISSON CALCULATIONS FOR SMALL MOLECULES: SOLVATION ENERGIES

FDPB calculations were first done with CHARMM22 charges and radii for the 13 solutes listed in Table I. Dependency on grid position is negligible. ACE calculations were performed with the same charges and radii. The atomic smoothing parameter α was first adjusted to optimize agreement between ACE and FDPB by using several volume assignment schemes. Agreement as a function of α is shown in Figure 1. The different volume schemes lead to approximately the same optimal $\alpha \approx 1.6$, which is in agreement with SK.[16] The solvation free energies are reported in Table I for four slightly different volume schemes. Only the continuum electrostatic contributions from ACE or FDPB are included, as discussed. With the OP Voronoi method for heavy atoms and zero volume for all hydrogens, the ACE energies have an rms error of only 0.86 kcal/mol compared to FDPB. The BP Voronoi method is very similar,

TABLE I.
ACE and FDPB Contribution to Solvation Free Energies (kcal / mol).

	FDPB		ACE			
			OP	OP	OP	BP
Solute	C22mod	C22	$r(H) = 0$	$r(H) = 1$	$r(H) = r_{vdW}$	$r(H) = 0$
Acetate ion	−87.66 (2)	−75.98 (2)	−75.88	−74.09	−72.56	−76.98
Alanine[a]	−13.52 (1)	−15.37 (3)	−13.88	−11.59	−9.86	−14.29
Imidazole[b]	−10.3 (1)	−10.86 (1)	−11.60	−10.56	−9.95	−11.54
Acetamide	−9.67 (1)	−9.12 (1)	−10.33	−9.19	−9.15	−10.59
Acetic acid		−7.38 (3)	−8.93	−8.36	−7.82	−9.15
Methanol	−7.38 (1)	−7.02 (1)	−7.89	−6.58	−5.78	−7.93
N-Methyl acetamide	−7.18 (1)	−6.94 (1)	−6.76	−5.86	−5.28	−6.9
Phenol	−6.48 (2)	−6.89 (3)	−5.93	−5.14	−4.38	−5.95
Ethanol	−6.03 (2)	−6.77 (0)	−6.85	−5.45	−4.45	−6.93
Propanol	−4.43 (3)	−6.32 (3)	−5.48	−3.97	−2.89	−5.79
Benzene	−2.31 (1)	−1.86 (1)	−1.59	−1.25	−0.85	−1.59
Butane	−0.10 (0)	−0.22 (0)	0.32	0.34	0.41	0.32
Ethane	−0.08 (0)	−0.17 (0)	0.07	0.11	0.20	0.07
Correlation[c]		1.00	0.92	0.92	0.92	0.92
RMSD		0.00	0.86	1.55	2.36	0.89
Mean dev.		0.00	0.70	1.19	1.81	0.74

Vapor to water transfer; standard states are 1 *M* in ideal solution and in the ideal gas phase. The solute dielectric constant is one. FDPB, finite-difference Poisson calculation (zero ionic strength). The standard deviation in significant digits in parentheses is from eight calculations with slightly different grid positions. C22mod, CHARMM22 radii modified by Nina et al.[30] C22, CHARMM22 charges and radii. Atomic volumes assigned from offset-plane (OP) or bisecting plane (BP) Voronoi method. The r is the radius used to assign hydrogen volumes ($= 4\pi r^3 / 3$ Å^3). RMSD, root mean square deviation from FDPB data. mean dev., mean absolute deviation from FDPB data.

[a] Alanine with acetylated N terminus and *N*-methylamide C terminus.

[b] Neutral, singly protonated form.

[c] Correlation coefficient with C22 FDPB data.

while larger hydrogen volumes lead to poorer agreement.

FDPB results with the radii optimized by Nina et al.[30] for Poisson calculations are also reported in Table I. Acetic acid was omitted from the calculations, because suitable atomic radii for the COOH could not be extrapolated from the amino acids studied in ref. 30. The FDPB results are significantly different for these radii. This sensitivity of FDPB to the force field is not of concern here. Our goal is to reproduce FDPB with ACE for a given force field; the optimization of the FDPB force fields themselves is beyond the scope of this article. Therefore, to reproduce the modified CHARMM22 FDPB data set, the same modified radii were used to perform the ACE calculations. The optimal α is again 1.6; however, the optimal volume scheme is now BP with hydrogen volumes fixed at $4\pi/3$. This illustrates the possibility of force field dependency of the optimal ACE protocol. The rms ACE error relative to FDPB is 1.48 kcal/mol (mean unsigned error 1.23 kcal/mol), which is somewhat larger than above.

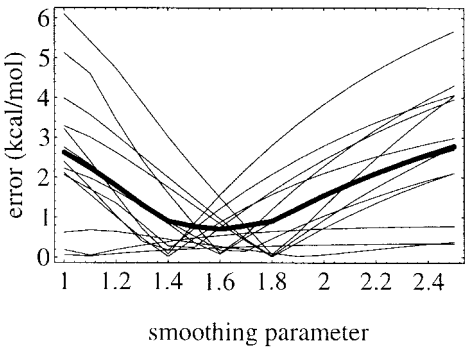

FIGURE 1. ACE performance as a function of the atomic smoothing parameter α for the small solutes listed in Table I. The thin lines are the unsigned difference between ACE and FDPB for individual solutes. The thick line is the mean unsigned difference averaged over all 13 solutes.

ACE calculations were also performed with OPLS charges and radii for the larger set of solutes listed in Table II. FDPB results from ref. 10, which were calculated with the Delphi program,[31] are reported for comparison. The optimal volume

TABLE II.
ACE and FDPB Contribution to Solvation Free Energies (kcal / mol).

	Solute	Area[a]	Exp[b]	FDPB		ACE[e]
				UHBD[c]	Delphi[d]	
1	Acetamide	216	−9.70	−10.7	−12.5	−13.6
2	Acetic acid	212	−6.73	−7.2	−8.6	−10.2
3	Methane thiol	192	−1.24	−3.9	−4.1	−6.6
4	Propionamide	247	−9.40	−11.1	−12.9	−13.3
5	Propionic acid	245	−6.47	−7.2	−8.6	−9.3
6	4-Methyl imidazole	259	−10.25	−8.8	−9.7	−9.7
7	3-Methyl indole	331	−5.90	−5.1	−6.2	−4.9
8	*p*-Cresol	303	−6.13	−6.6	−7.9	−5.3
9	Ethyl methyl sulfide	246	−1.49	−3.1	−3.3	−3.0
10	Methanol	174	−5.10	−6.1	−7.2	−8.9
11	Ethanol	206	−4.95	−5.8	−6.9	−7.5
12	Toluene	297	−0.76	−1.3	−1.8	−0.9
13	Butyl amine	272	−4.38	−2.4	−3.4	−1.1
14	Ethane thiol	220	−1.30	−3.7	−4.4	−5.9
15	Dimethyl sulfide	222	−1.54	−3.1	−3.5	−3.3
16	*N*-Methyl acetamide	254	−10.07	−8.1	−9.6	−9.4
17	Acetone	233	−3.80	−3.6	−4.0	−3.5
18	Butanoic acid	274	−6.35	−7.2	−9.2	−9.3
19	1-Propanol	236	−4.84	−5.9	−7.0	−6.7
20	2-Propanol	235	−4.75	−5.7	−6.6	−6.3
21	2-Methyl-2-propanol	260	−4.52	−4.9	−5.8	−5.2
22	Methyl acetate	246	−3.31	−2.2	−2.8	−2.3
23	Ethyl acetate	280	−3.09	−3.1	−3.7	−2.7
24	Methyl amine	178	−4.57	−4.1	−5.0	−4.0
25	Ethyl amine	209	−4.56	−3.8	−4.7	−3.3
26	Propyl amine	241	−4.45	−3.8	−4.7	−3.0
27	Benzene	275	−0.88	−2.7	−3.7	−2.4
28	Phenol	253	−6.50	−6.6	−8.4	−5.4
29	Dimethyl acetamide	272	−8.54	−5.6	−6.4	−5.1
RMS error[f]						1.33

OPLS charges and radii (with selected hydrogen radii set equal to 1 Å.[5,10]) Vapor to water transfer. Standard states are 1 *M* in ideal solution and in the ideal gas phase. The solute dielectric constant is one. Experimental data are reported, but they can only be qualitatively compared to ACE and FDPB, which only include electrostatic contributions.

[a] Solvent accessible surface area (Å^2).

[b] Experimental solvation free energies. See references in ref. 10.

[c] Calculated with the UHBD program,[36] which uses a smoothed dielectric boundary.

[d] Calculated with the Delphi program (version 3.0, Kim Sharp, 1988),[31] which uses an abrupt solute–solvent boundary. Values taken from ref.[10].

[e] OP volume method, zero hydrogen volumes, smoothing parameter $\alpha = 1.6$.

[f] Relative to Delphi.

scheme is the same as above (OP with zero hydrogen volumes), and the optimal α is again 1.6. Thus, the ACE protocol used with the CHARMM22 force field is transferable to the OPLS force field. The ACE rms error relative to FDPB is 1.33 kcal/mol (mean unsigned error 1.07 kcal/mol). FDPB values calculated with UHBD are also reported in the Table. The differences relative to Delphi are of the same order as the differences between Delphi and ACE. This is presumably a consequence of the different dielectric boundary definitions in the two programs. Although both programs use the molecular surface to define the boundary, UHBD uses a boundary smoothing method[28,29] that effectively blurs the dielectric boundary slightly, while the version 3.0 of Delphi used in ref. 10 assumes an abrupt boundary. Similar observations have been reported by others.[32]

Below, the Gaussian nonpolar term will be combined with the present FDPB electrostatic contri-

bution and their sum compared to experimental solvation free energies.

POISSON CALCULATIONS FOR SMALL MOLECULES: CONFORMATIONAL ENERGIES

To test the ability of ACE to discriminate between different conformations of a single molecule, we calculated the energies of aspartate and asparagine as a function of their side chain torsional angles χ_1 and χ_2. Each amino acid has three stable free energy minima around χ_1, which are broad and relatively flat in the χ_2 direction.[24] Energy maps are shown in Figure 2 that are calculated with either the ACE implicit solvent model or a uniform dielectric constant of 20. These solvent models are combined with the CHARMM22 internal energy. Contributions from nonelectrostatic solute-solvent interactions are expected to approximately cancel when different conformations are compared (see below). Points sampled in molecular dynamics simulations in explicit water[24] are overlaid for comparison. The conformational energies are compared in Table III with results from free energy simulations in explicit water [molecular dynamics-free-energy perturbation (MD-FEP) entry in Table III].[24] The MD-FEP result $\Delta G(\text{Asp} \rightarrow \text{Asn})$ corresponds to the alchemical transformation of Asp into Asn in water. For comparison, the Asp → Asn ACE or FDPB solvation free energy difference must be added to the free energy for transforming Asp into Asn *in vacuo*. The latter was estimated from molecular dynamics in the gas phase (G. Archontis, T. Simonson, and M. Karplus, unpublished data) to be +22.9 kcal/mol.

The (χ_1, χ_2) energy maps with $\varepsilon = 20$ and ACE are qualitatively similar, and the shape of the energy basins qualitatively agrees with the explicit solvent simulations that explored them. The relative energies of the stable conformers agree only roughly with the MD-FEP calculations, and the gauche conformers are too high compared to the explicit solvent (Table III).

The Asp → Asn free energy difference is underestimated by ACE (68 vs. 80 kcal/mol from MD-FEP). Detailed FDPB calculations give significantly better agreement with MD-FEP [$\Delta G(\text{Asp} \rightarrow \text{Asn}) = 75$ kcal/mol]. The use of molecular dynamics data for the *in vacuo* step of the calculation is critical to the success of the FDPB (and ACE) calculations. Indeed, if single point vacuum energies are used, the result is very dependent on the single conformation chosen and a large uncertainty ensues.

POISSON CALCULATIONS FOR PROTEINS: SOLVATION ENERGIES

Solvation energies were calculated for BPTI, crambin, and Ras. Structures were extracted from the Protein Data Bank, and hydrogens were constructed automatically. Guanosine tri-phosphate and magnesium bound to Ras were not included in the calculations. CHARMM19 charges and radii were used; the protein and solvent dielectric constants were 1 and 80, respectively. Atomic volumes for ACE were mainly taken from the library of ref. 16, which assigns a zero volume to hydrogens. Volume calculations with Voronoi polyhedra were performed for comparison (results not shown). Hydrogens were normally excluded from these calculations and assigned zero volume; when included, the volume construction routinely failed for certain hydrogen atoms. A significant variability of volumes was observed among atoms having the same chemical type (as defined, for example, in the SK library[16]), which was due to differences in their environment (see below). Although ACE results were fairly robust with respect to these variations, calculations reported here used the volume library.

ACE and FDPB results are compared in Table IV. FDPB results with grid spacings of 0.4 and 0.5 Å agree, indicating that convergence with respect to grid spacing is satisfactory and sensitivity to grid position is very weak. ACE results with several volume schemes are reported; the optimal scheme uses OP with zero hydrogen volumes as above. The agreement with FDPB is good, provided α is optimized separately for each protein. Thus, Ras and crambin give good agreement with $\alpha = 1.2$ and fair agreement with $\alpha = 2.5$, while BPTI requires $\alpha \approx 2.5$. The value of 1.9 suggested by SK[16] gives only qualitative agreement for the present cases. More recently, good results were obtained in other systems by choosing α to minimize density fluctuations within the solute (M. Schaefer, personal communication).

POISSON CALCULATIONS FOR PROTEINS: Ras CONFORMATIONAL ENERGIES

An important potential application of implicit solvent models is the simulation of flexible surface regions of proteins, such as the loop residues 30–37 in the effector domain of Ras. Seventy-four sterically favorable conformations of this loop, which were generated by systematic searching and conformational clustering (J. Zeng, unpublished data),

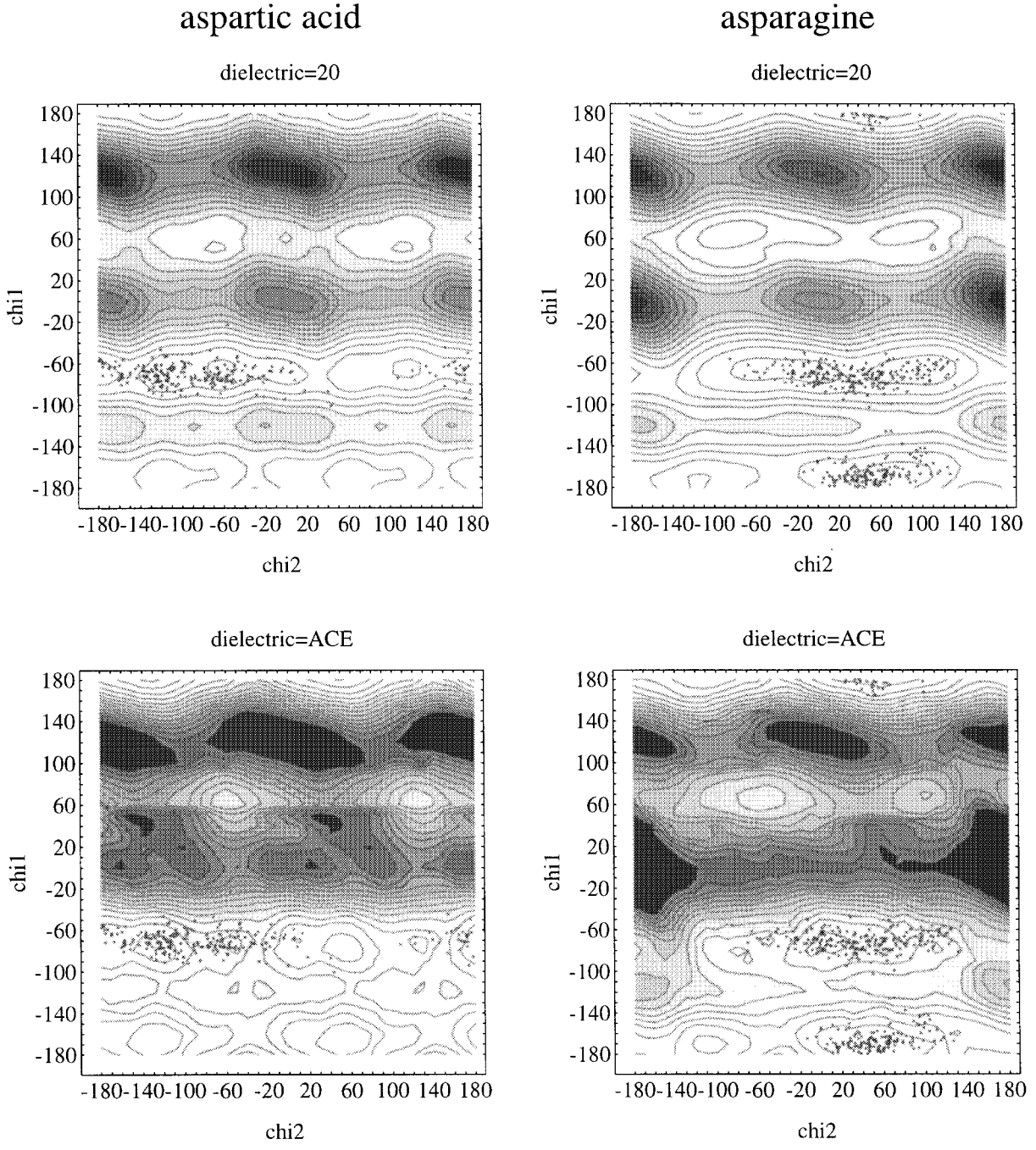

FIGURE 2. Adiabatic energy maps for the amino acids Asp (left) and Asn (right) as a function of their side chain dihedral angles. The solvent was modeled with a uniform dielectric constant of 20 (above) or with ACE (below). Points from molecular dynamics simulations in explicit solvent are shown as dots. Low energy regions are light; high energy regions are dark. Contour levels are separated by 1 kcal / mol.

were subjected to FDPB and ACE calculations. A subset of the conformations are shown schematically in Figure 3a. Only protein residues within 8 Å of the loop residues were retained in the calculations, and atomic partial charges were set to zero except for residues 28–42.

Volumes were assigned either from the SK library[16] or from Voronoi polyhedra. Variations of the Voronoi volumes among atoms having the same chemical type are summarized in Table V. Carbon volumes throughout the loop vary by 3.1–3.5 Å^3 for a given chemical type and a given

TABLE III.
ACE and MD-FEP Conformational Energies (kcal / mol).

Solute	Conformation	Free Energy[a]		Population[b] (%)	
		MD-FEP[c]	ACE	MD	ACE
Asp	Gauche +	2.0	5.2	2	0
Asp	Gauche −	0	1.5	49	7
Asp	Trans	0	0	49	93
Asn	Gauche +	2.0	5.6	2	0
Asn	Gauche −	0	2.7	49	1
Asn	Trans	0	0	49	99
ΔG(Asp → Asn)[d]		79.8	68.0[e]		

[a] Relative to the trans conformer.
[b] Estimated from Boltzmann factors.
[c] Conformational free energies from explicit solvent molecular dynamics simulations.[24]
[d] Free energy to mutate Asp into Asn in solution. MD-FEP value from ref 24.
[e] The ACE value includes a +22.9 kcal / mol contribution corresponding to the Asp → Asn mutation *in vacuo* (see text).

conformation and by 0.4–2.4 Å^3 for a given atom when different loop conformations are considered. Nitrogen and oxygen volumes vary less. The latter are much smaller than the SK library values,[16] but this is largely compensated for by increased carbon volumes. For example, while the carbonyl oxygen has a 4.3 Å^3 Voronoi volume and a 15.6 Å^3 SK volume,[16] the Voronoi volume of a CO group is 20.9 Å^3 on average in the loop compared to 24.3 Å^3 in the SK library.[16] In addition, the ACE results were quite robust with respect to these variations [e.g., the loop energies changed by only 4 kcal/mol (rms difference) when Voronoi volumes were used instead of SK volumes].

The rms deviation between the FDPB and ACE energies with an optimal $\alpha = 2$ and SK[16] volumes is 6.96 kcal/mol, which is just 1% of the average total energy or 7% of the energy difference between the lowest and highest energy conformations (Fig. 3b). For the pairwise energy differences between conformations, the rms error is 9.34 kcal/mol with an optimal α of 2.1. Because 9.34 is only slightly less than $6.96\sqrt{2} = 9.84$, we may conclude that only a slight cancellation of error occurs for the energy differences. The rank correlation coefficient of 0.93 and the comparison of energies in Figure 3b show that the FDPB ordering of conformations is well reproduced by ACE with some differences, though, in the important low energy region. Results with Voronoi volumes are also shown in the figure and are remarkably similar (rms error of 6.67 kcal/mol compared to FDPB, optimal $\alpha = 1.95$), despite the significant differences in atomic volumes.

For this intermediate-sized system (on a workstation with 768 megabytes of memory) a single energy calculation with ACE was 700 times faster than with FDPB. Memory requirements for ACE are very low, but a large memory enhances the FDPB performance.

GAUSSIAN NONPOLAR MODEL

A comprehensive implicit solvent model requires that the continuum electrostatic term (FDPB or ACE) be augmented by a nonelectrostatic or nonpolar term.[6–10] To this end, the Gaussian solvation model of Lazaridis and Karplus[19,20] was used to calculate the solvation free energies of 23 saturated alkanes, including seven cyclic alkanes. The model contains three alkane atom types. The alkane parameters of Lazaridis and Karplus[19,20] were modified by applying an overall scaling factor of 1.24 to the reference energies and changing λ from 3.5 to 5.75; the original and modified parameters are reported in Table VI. The experimental free energies are then fit with an rms error of 0.19 kcal/mol (Table VII). The same parameters

TABLE IV.
ACE and FDPB Protein Solvation Free Energies (kcal / mol).

Protein	FDPB		ACE				
	GS = 0.5 Å	GS = 0.4 Å	$\alpha = 1.2$	$\alpha = 1.5$	$\alpha = 1.9$	$\alpha = 2.2$	$\alpha = 2.5$
BPTI	−1611 (3)	−1602	−1199	−1410	−1504	−1538	−1581
Crambin	−344 (1)	−342	−343	−344	−285	−281	−304
Ras	−2497 (1)	−2486	−2383	−3039	−3160	−2905	−2695

Protein dielectric constant of one. GS, final grid spacing. α, ACE smoothing parameter. The values in parentheses are the standard deviation in significant digits from eight runs with different grid positions (only available for the 0.5-Å data).

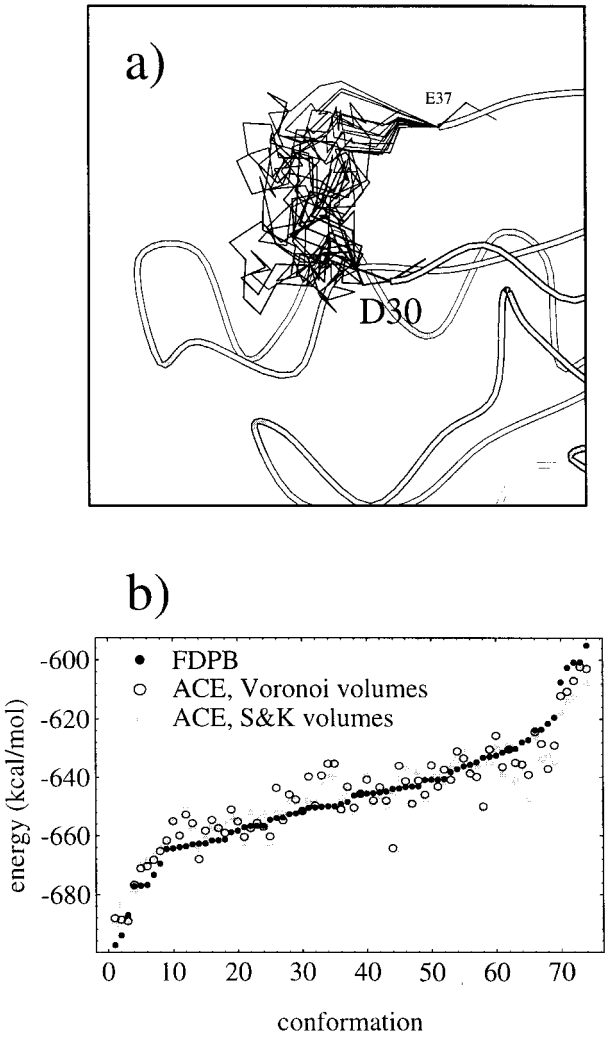

FIGURE 3. (a) Detail of the backbone of the protein Ras, including 16 conformations of loop residues 30–37, out of the 74 conformations used in the calculations (see text). Resides 30–37 are shown in stick representation; the rest of the backbone is shown in tube form. The figure was produced with the Molscript program.[37] (b) Solvation energies of the 74 Ras loop conformations.

are used for cyclic and noncyclic compounds. This contrasts with a surface area model which requires only one surface tension for linear and moderately branched alkanes but requires at least three distinct surface tensions to fit cyclic alkane data with the same accuracy.[10] This suggests that the Gaussian model may be more robust with respect to the environment of each group and more transferable to a protein environment.

Preliminary attempts were made to fit the solvation free energies of the polar compounds in Table II by combining the FDPB electrostatic term with the Gaussian nonpolar term. CHARMM19 atomic radii were used, along with the atomic volumes and the single λ of Lazaridis and Karplus.[19,20] Three parameters were adjusted: single values of the two reference energies G_i^{ref} and G_i^{free} that are applicable to all non-alkane atoms and a scaling factor applied uniformly to the reference energies of all alkane atoms. The agreement for the total solvation free energies (not shown) was reasonable (1.2 kcal/mol RMS errors). However, if the FDPB term was subtracted from the experimental solvation free energy to give an ''inferred'' nonpolar term, its correlation with the calculated nonpolar term was only 46%. Although this is double the correlation obtained with a simple surface area model,[10] a more sophisticated adjustment of parameters is presumably required to combine the electrostatic and nonpolar terms into a fully satisfactory hybrid model.

Concluding Discussion

Continuum electrostatics is known to provide a useful implicit solvent model, which is capable of reproducing a variety of thermodynamic properties of small solutes and proteins.[1–5,33] Therefore, analytical approximations to continuum electrostatics, which have high computational efficiency, are also attractive alternatives to explicit solvent models. They can be combined with simple models of nonpolar interactions, such as surface area models or the Gaussian model of Lazaridis and Karplus.[19,20] To fully test their usefulness they must be compared to numerical solutions of the Poisson equation, explicit solvent simulations, and experimental data.

We focus on the ACE approach of SK.[16] The ACE solvation free energies were primarily compared to FDPB. FDPB in turn is shown to give reasonable accuracy for the electrostatic component of solvation free energies, compared to experiment[9,10] and explicit solvent simulations.[4,5,30] The ACE results were in reasonable agreement with FDPB for small solutes and proteins with several force fields, provided the ACE protocol (i.e., volume assignment, smoothing parameter, treatment of hydrogens) was adjusted for each force field and family of solutes. If atomic volumes were assigned from an existing library and the smoothing parameter was chosen to minimize density fluctuations in the solute (M. Schaefer, personal communication), no protocol adjustment

TABLE V.
Atomic Volumes for Ras Loop.

Atom Type	CHARMM19 Code	No. Occurrences[a]	Overall Average[b]	Single Conf. SD[c]	SD between Confs.[d]	SK[e]
Carbonyl C	C	15	16.6	4.5	0.8	8.7
Alkane CH	CH1E	11	29.8	4.1	1.9	15.4
Alkane CH_2	CH2E	11	28.3	5.5	2.4	25.5
Alkane CH_3	CH3E	3	33.7	3.1	0.4	39.0
Ring CH	CR1E	4	23.0	3.3	0.3	18.8–22.1
Amide N	NH1	8	3.3	0.5	0.2	13.4
Carbonyl O	O	9	4.3	0.5	0.2	15.6
Carboxylate O	OC	8	6.7	0.8	0.4	21.1

Volumes (Å^3) for the most frequent CHARMM19 atom types within the 30–37 loop residues.
[a]Number of atoms of the given type within the loop.
[b]Volume averaged over all atoms of the given type and all 74 loop conformations.
[c]Standard deviation when different atoms having the same chemical type are considered for a single loop conformation. (For simplicity, results averaged over the 74 loop conformations are presented.)
[d]Standard deviation of the individual atomic volumes when different loop conformations are considered. (Results averaged over all atoms of a given chemical type are presented.)
[e]SK library value(s).[16] The library distinguishes three types of ring carbons with the range of volumes indicated.

was necessary. However, until more experience is gained, it is probably preferable to explore a range of volume assignments and smoothing parameters for each system, as in this work. Such adjustments are comparable to those of FDPB protocols for specific systems where the solute–solvent interface, solvent probe radius, and overall boundary conditions must all be suitably chosen. As observed by others,[32] significant differences are found between FDPB calculations that use slightly different definitions of the solute–solvent interface; these differences are of the same magnitude as the differences between ACE and FDPB and between FDPB and experiment. This underscores the need to calibrate ACE and FDPB for each family of systems and each force field. Optimization of the force field itself (i.e., atomic charges and radii) is a much vaster problem that was not considered here but is common to ACE and FDPB.

The ACE solvation free energies of a range of small solutes are in good agreement with FDPB for several choices of force field. The exact Voronoi algorithm used is not critical, but the assignment of zero volume to hydrogens is strongly preferred. This is consistent with the assignment of zero van der Waals radii to polar hydrogens in the force fields considered (CHARMM22 and OPLS).

The ACE conformational energies of two amino acids, the neutral asparagine and the negatively charged aspartate, were compared to FDPB and explicit solvent simulations. It is important to compare to the latter, because the accuracy of FDPB for this kind of application is less well established than for solvation free energies. For example, only

TABLE VI.
Gaussian Solvation Model Parameters.

	Atom Type	Volume (Å^3)	R_{vdw} (Å)	Original[a] (kcal / mol)		Adjusted[b] (kcal / mol)	
				G_i^{ref}	G_i^{free}	G^{ref}	G^{free}
CH	CH1E	23.7	2.365	−0.187	−0.25	−0.231	−0.31
CH_2	CH2E	22.4	2.235	0.372	0.52	0.460	0.64
CH_3	CH3E	30.0	2.165	1.089	1.50	1.346	1.85

Atom types refer to the CHARMM19 force field. The van der Waals radii are from the CHARMM19 force field (position of the minimum in the van der Waals potential).
[a]Original parameters of Lazaridis and Karplus.[19,20]
[b]Adjusted to fit alkane data by applying a scaling factor of 1.24.

TABLE VII.
Alkane Solvation Free Energies (kcal / mol).

	Solute	Area[a]	Exp	Gaussian Model	γA
1	Ethane	182	1.77	1.96	1.09
2	Propane	223	1.98	2.10	1.34
3	*n*-Butane	250	2.15	2.22	1.50
4	Isobutane	250	2.28	2.31	1.50
5	*n*-Pentane	275	2.34	2.26	1.65
6	Isopentane	275	2.38	2.39	1.65
7	neopentane	270	2.68	2.57	1.62
8	*n*-Hexane	301	2.55	2.36	1.81
9	Isohexane	294	2.53	2.48	1.76
10	3-Methyl pentane	299	2.51	2.52	1.79
11	Neohexane	291	2.60	2.65	1.75
12	*n*-Heptane	332	2.63	2.53	1.99
13	2-4-Dimethyl pentane	326	2.89	2.80	1.96
14	*n*-Octane	365	2.87	2.68	2.19
15	Isooctane	341	2.85	2.94	2.05
16	2-2-5-Trimethyl hexane	371	2.72	3.05	2.23
17	Cyclopentane	241	1.21	0.97	1.45
18	Cyclohexane	260	1.24	1.04	1.56
19	Methyl-cyclopentane	275	1.60	1.36	1.65
20	Methyl-cyclohexane	291	1.69	1.44	1.75
21	Cycloheptane	282	0.80	1.14	1.69
22	Cyclooctane	305	0.84	1.24	1.83
23	cis-1-2-Dimethyl cyclohexane	209	1.57	1.76	1.85
RMS error				0.19	0.69
Mean unsigned error				0.16	0.64

Standard states are 1 *M* in ideal solution and in the ideal gas phase. For experimental values see the references in ref. 10. Calculated values are from the Gaussian nonpolar solvation model.
[a] Solvent accessible surface area (Å^2).

qualitative agreement is seen between FDPB and explicit solvent simulations for the potential of mean force between small amide pairs.[10,34] Here the ACE results are in only qualitative agreement with explicit solvent simulations (i.e., the energy of the gauche conformations is overestimated by 2–4 kcal/mol). ACE contains only electrostatic contributions, in contrast to the explicit solvent simulations. However, the accessible surface areas of the different Asp and Asn conformers are very similar, so that if we assume nonpolar terms can be approximated by an accessible surface area treatment, the free energy differences between conformation should be dominated by the electrostatic term. Another limitation of the ACE calculation performed here is that a single structure was used for each conformation (i.e., intramolecular entropy contributions were neglected). A rigorous ACE calculation would require a molecular dynamics or Monte Carlo free energy perturbation approach with ACE as a solvent model.

The free energy difference between Asp and Asn is also reproduced qualitatively by ACE. Because Asp and Asn are almost isosteric, the nonelectrostatic contributions are expected to be small for the same reason as above. Indeed, the free energy difference from explicit solvent simulation is dominated by electrostatics.[24] The ability of ACE to distinguish at least qualitatively between a charged and a neutral amino acid and their important conformers suggests that in a protein context ACE will approximately respect the balance of interactions of charged and polar amino acids among themselves and with solvent.

Two applications to proteins were considered: absolute solvation free energy calculations and conformational free energy calculations. In each case, ACE was compared to FDPB alone. Direct comparison to an experiment is difficult because of limited available data and the presence of nonelectrostatic contributions. Data from explicit solvent simulation is also very limited. Thus, futher com-

parisons are needed and will be the object of additional studies. Absolute protein solvation free energies are difficult to calculate by theoretical methods and are important for studies of protein stability and protein–protein binding. Good agreement between ACE and FDPB for BPTI, crambin, and Ras was obtained, albeit with three slightly different ACE protocols. As with the small solutes, a precalibration of ACE is thus required for each application, which can be achieved at a low computational cost through selected FDPB calculations.

Applications of implicit solvent models to protein conformational free energies are especially important. Theoretical studies of protein thermodynamics are plagued by the multiple minimum problem (see, e.g., ref. 35 and references therein), and fast and reliable methods to rank alternate conformations in order of decreasing statistical weight are sorely needed. FDPB models are thought to achieve this at least roughly, so that the good agreement between ACE and FDPB for the ranking of 74 Ras loop conformations is very promising. The possibility that the optimal ACE protocol could vary with conformation was not explored in detail; however, atomic volume fluctuations between conformations were moderate (Table V). Variations in the optimal smoothing parameter are possible in principle; calculations with several values would be needed for at least a few conformations to test this in detail. However, the good agreement with FDPB suggests that this effect is not very important here. Thus, if ACE were used to select the lowest energy loop conformations (say, for further study with an explicit solvent model) eight of the first 10 conformations selected would be ''correct'' (i.e., they would also have been selected by FDPB) and 18 of the first 25 (which span more than 40 kcal/mol in energy). As with the small solute solvation free energies, the discrepancies between ACE and FDPB for this application are probably smaller than the uncertainty of FDPB itself.

A complete implicit solvent model requires, at the least, a simple model of nonpolar interactions in addition to the continuum electrostatics term. Accessible surface models were considered in the past with moderate success, and we presented alkane solvation results using the Gaussian model of Lazaridis and Karplus.[19,20] Because the model is already parameterized to reproduce solvation free energies of small polar and nonpolar solutes, it is not surprising that only slight reparameterization was needed to obtain very good agreement for 23 alkane solvation free energies (Table VII). The combination of this model with an FDPB treatment of electrostatic interactions is a more complex problem, and our preliminary results suggest that a more sophisticated and extensive reparameterization of the Gaussian term will be needed.

Efficient inclusion of ACE in molecular dynamics programs is complicated by the presence of three-body forces.[16] Furthermore, ACE can be implemented alone or in combination with a nonpolar or ''hydrophobic'' energy term. Applications to explore these possible implementations are under way.

Acknowledgments

We thank Jun Zeng for providing the Ras loop conformations, Michael Schaefer and Themis Lazaridis for discussions, Themis Lazaridis and Martin Kaplus for providing data prior to publication, and Mafalda Nina for comments on the manuscript.

References

1. Roux, B.; Simonson, T., Eds. Biophys Chem (Special Issue) to appear.
2. Kirkwood, J.; Westheimer, F. J Chem Phys 1938, 6, 506–512.
3. Lim, C.; Bashford, D.; Karplus, M. J Phys Chem 1991, 95, 5610–5620.
4. Jean–Charles, A.; Nicholls, A.; Sharp, K.; Honig, B.; Tempzyck, A.; Hendrickson, T.; Still, W. J Am Chem Soc 1991, 113, 1454–1455.
5. Mohan, V.; Davis, M.; McCammon, J. A.; Pettitt, B. M. J Phys Chem 1992, 96, 6428–6431.
6. Huron, M.; Claverie, P. J Phys Chem 1972, 76, 2123–2133.
7. Cramer, C.; Truhlar, D. Science 1992, 256, 213–217.
8. Hawkins, G.; Cramer, C.; Truhlar, D. J Phys Chem B 1997, 101, 7147–7157.
9. Sitkoff, D.; Sharp, K.; Honig, B. J Phys Chem 1994, 98, 1978–1988.
10. Simonson, T.; Brünger, A. T. J Phys Chem 1994, 98, 4683–4694.
11. Gilson, M.; Davis, M.; Luty, B., McCammon, J. J Phys Chem 1993, 97, 3591–3600.
12. Still, W. C.; Tempczyk, A.; Hawley, R.; Hendrickson, T. J Am Chem Soc 1990, 112, 6127–6129.
13. Sklenar, H.; Eisenhaber, F.; Poncin, M.; Lavery, R. In Theoretical Biochemistry & Molecular Biophysics; Beveridge, D.; Lavery, R., Eds.; Adenine Press: Schenectady, NY, 1990; pp 317–335.
14. Schaefer, M.; Froemmel, C. J Mol Biol 1990, 216, 1045–1066.
15. Hawkins, G.; Cramer, C.; Truhlar, D. Chem Phys Lett 1995, 246, 122.
16. Schaefer, M.; Karplus, M. J Phys Chem 1996, 100, 1578–1599.

17. Horvath, D.; van Belle, D.; Lippens, G.; Wodak S. J Chem Phys 1996, 104, 6679–6695.
18. Hawkins, G.; Cramer, C., Truhlar, D. J Phys Chem 1996, 100, 19824.
19. Lazaridis, T.; Karplus, M. Science 1997, 278, 1928–1931.
20. Lazaridis, T.; Karplus, M. Proteins to appear.
21. Jorgensen, W.; Tirado–Rives, J. J Am Chem Soc 1988, 110, 1657–1666.
22. Mackerell, A.; Bashford, D.; Bellott, M.; Dunbrack, R.; Evanseck, J.; Field, M.; Fischer, S.; Gao, J.; Guo, H.; Ha, S., Joseph, D.; Kuchnir, L.; Kuczera, K.; Lau, F.; Mattos, C.; Michnick, S.; Ngo, T., Nguyen, D.; Prodhom, B.; Reiher, W.; Roux, B. Smith, J.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz–Kuczera, J.; Yin, D.; Karplus, M. J Phys Chem B 1998, 102, 3586–3616.
23. Brooks, B.; Bruccoleri, R.; Olafson, B.; States, D.; Swaminathan, S.; Karplus, M. J Comput Chem 1983, 4, 187–217.
24. Archontis, G.; Simonson, T.; Moras, D.; Karplus, M. J Mol Biol 1998, 275, 823–846.
25. Lee, B.; Richards, F. J Mol Biol 1971, 55, 379–400.
26. Brünger, A. T. X-PLOR Version 3.1, A System for X-Ray Crystallography and NMR. Yale University Press; New Haven, CT, 1992.
27. Richards, F. VOLUME Program. Yale University Press: New Haven, CT, 1983.
28. Davis, M.; Madura, J.; Luty, B.; McCammon, J. Comput Phys Commun 1991, 62, 187–197.
29. Madura, J.; Briggs, J.; Wade, R.; Davis, M.; Luty, B.; Ilin, A.; Antosiewicz, J.; Gilson, M.; Baheri, B.; Scott, L.; McCammon, J. Comput Phys Commun 1995, 91, 57–95.
30. Nina, M.; Beglov, D.; Roux, B. J Phys Chem 1997, 101, 5239–5248.
31. Nicholls, A.; Honig, B. J Comput Chem 1991, 12, 435–445.
32. Scarsi, M.; Apostolakis, J.; Caflisch, A. J Phys Chem A 1997, 101, 8098–8106.
33. Honig, B.; Nicholls, A. Science 1995, 268, 1144.
34. BenTal, N.; Sitkoff, D.; Topol, I.; Yang, A.; Burt, S.; Honig, B. J Phys Chem B 1997, 101, 450–457.
35. Hodel, A.; Rice, L.; Simonson, T.; Fox, R. O.; Brünger, A. T. Protein Sci 1995, 4, 634–654.
36. Davis, M.; Madura, J.; Luty, B.; McCammon, J. A. Comput Phys Commun 1991, 62, 187.
37. Kraulis, P. J Appl Crystallogr 1991, 24, 946–950.